REGULARITY OF SOLUTIONS OF THE STEIN EQUATION AND RATES IN THE MULTIVARIATE CENTRAL LIMIT THEOREM

T. O. GALLOUET, G. MIJOULE, AND Y. SWAN

Abstract. Consider the multivariate Stein equation $\Delta f - x \cdot \nabla f = h(x) - \mathbb{E}h(Z)$, where Z is a standard d-dimensional Gaussian random vector, and let f_h be the solution given by Barbour's generator approach. We prove that, when h is α-Hölder ($0 < \alpha \leqslant 1$), all derivatives of order 2 of f_h are α-Hölder *up to a* log *factor*; in particular they are β-Hölder for all $\beta \in (0, \alpha)$, hereby improving existing regularity results on the solution of the multivariate Gaussian Stein equation. For $\alpha = 1$, the regularity we obtain is optimal, as shown by an example given by Raič [18]. As an application, we prove a near-optimal Berry-Esseen bound of the order $\log n/\sqrt{n}$ in the classical multivariate CLT in 1-Wasserstein distance, as long as the underlying random variables have finite moment of order 3. When only a finite moment of order $2+\delta$ is assumed ($0 < \delta < 1$), we obtain the optimal rate in $\mathcal{O}(n^{-\frac{\delta}{2}})$. All constants are explicit and their dependence on the dimension d is studied when d is large.

Keywords. Berry-esseen bounds; Stein's method; Elliptic regularity;

AMS subjects classification.

1. Introduction

1.1. **Multivariate Stein's method.** Stein's method is a powerful tool for estimating distances between probability distributions. It first appeared in [22], where the method was introduced for the purpose of comparison with a (univariate) Gaussian target. The idea, which still provides the backbone for the contemporary instantiations of the method, is as follows. If Z is a standard Gaussian random variable, then

$$\mathbb{E}[f'(Z) - Zf(Z)] = 0 \tag{1}$$

for all absolutely continuous functions f with $\mathbb{E}|f'(Z)| < +\infty$. Let X be another random variable, and consider the integral probability distance between the laws of X and Z given by

$$d_{\mathcal{H}}(X, Z) = \sup_{h \in \mathcal{H}} \mathbb{E}[h(X) - h(Z)], \tag{2}$$

for $\mathcal{H}$ a class of tests functions which are integrable with respect to the laws of both X and Z. Many classical distances admit a representation of the form (2), including the Kolmogorov (with $\mathcal{H}$ the characteristic functions of half-lines), total variation (with $\mathcal{H}$ the characteristic functions of Borel sets), and 1-Wasserstein a.k.a. Kantorovitch (with $\mathcal{H}$ the 1-Lipschitz real functions) distances; see e.g. [15]. Letting ω denote the standard Gaussian pdf, we define for every $h \in \mathcal{H}$ the function

$$f_h(x) = \frac{1}{\omega(x)} \int_{-\infty}^{x} (h(y) - \mathbb{E}h(Z))\, \omega(y) dy. \tag{3}$$

This function is a solution to the ODE (called a *Stein equation*)

$$f_h'(x) - x f_h(x) = h(x) - \mathbb{E}h(Z), \quad x \in \mathbb{R}, \tag{4}$$

which allows to rewrite integral probability metrics (2) as

$$d_{\mathcal{H}}(X, Z) = \sup_{h \in \mathcal{H}} \mathbb{E}[f_h'(X) - X f_h(X)].$$

Stein's intuition was to exploit this last identity to estimate the distance between the laws of X and Z. One of the reasons for which the method works is the fact that the function f_h defined in (3) enjoys many regularity properties. For instance, one can show (see e.g. [17, pp. 65-67]) that if h is absolutely continuous then

$$\|f_h\|_\infty \leq 2\|h'\|_\infty, \quad \|f_h'\|_\infty \leq \sqrt{2/\pi}\|h'\|_\infty \text{ and } \|f_h''\|_\infty \leq 2\|h'\|_\infty, \tag{5}$$

$\|\,.\,\|_\infty$ holding for the supremum norm. This offers a wide variety of handles on $d_{\mathcal{H}}(X,Z)$ – typically via low order Taylor expansion arguments – for all important choices of test functions $\mathcal{H}$ and under weak assumptions on X. This observation has been used, for instance, to obtain Berry-Esseen-type bounds in the classical central limit theorem in 1-Wasserstein distance, Kolmogorov or total variation distances, see [8, 17].

Consider now a d-dimensional Gaussian target $Z \sim \mathcal{N}(0, I_d)$. The d-dimensional equivalent to identity (1) was identified in [1, 13] as

$$\mathbb{E}[\Delta f(Z) - Z \cdot \nabla f(Z)] = 0,$$

which holds for a "large class" of functions $f : \mathbb{R}^d \to \mathbb{R}$ ($x \cdot y$ denotes the usual scalar product between vectors $x, y \in \mathbb{R}^d$). We will define the "large class" of functions precisely in Proposition 2.1 below. For h a function with finite Gaussian mean, the multivariate Stein equation then reads

$$\Delta f(x) - x \cdot \nabla f(x) = h(x) - \mathbb{E}h(Z), \quad x \in \mathbb{R}^d. \tag{6}$$

Note that (6) is a second order equation in the unknown function f; in dimension $d = 1$, (6) reduces to $f''(x) - xf'(x) = h(x) - \mathbb{E}h(Z)$ which is obtained by applying (4) to f'. Barbour [1] identified a solution of (6) to be

$$f_h(x) = -\int_0^1 \frac{1}{2t}\mathbb{E}[h(\sqrt{t}x + \sqrt{1-t}Z) - h(Z)]\,dt, \tag{7}$$

and the same argument as in the 1-dimensional setting leads to the identity

$$d_{\mathcal{H}}(X,Z) = \sup_{h \in \mathcal{H}} \mathbb{E}\left[\Delta f_h(X) - X \cdot \nabla f_h(X)\right], \tag{8}$$

which is the starting point for multivariate Gaussian approximation via Stein's method. The explicit representation (7) is suitable to obtain regularity properties of f_h in terms of those of h; for instance (see e.g. [19, Lemma 2.6]) it is known that if h is n times differentiable then f_h is n times differentiable and

$$\left|\frac{\partial^k f_h(x)}{\prod_{j=1}^k \partial x_{i_j}}\right| \leq \frac{1}{k}\left|\frac{\partial^k h(x)}{\prod_{j=1}^k \partial x_{i_j}}\right|, \tag{9}$$

for every $x \in \mathbb{R}^d$. Hence, contrarily to the univariate case where first order assumption on h was sufficient to deduce second order regularity for f_h (recall (5)), a bound such as (9) only shows the same regularity for h and f_h. In most practical implementations of the method, however, Taylor expansion-type arguments are used to obtain the convergence rates from the rhs of (8); hence regularity of f_h is necessary in order for the argument to work. This restricts the choice of class $\mathcal{H}$ in which the statements are made and therefore weakens the strength – be it only in terms of the choice of distance – of the resulting statements.

An important improvement in this regard is due to Chatterjee and Meckes [7] who obtained (among other regularity results) that

$$\sup_{x \in \mathbb{R}^d} \left\|\nabla^2 f_h(x)\right\|_{H.S} \leqslant \|\nabla h\|_\infty, \tag{10}$$

($\|M\|_{H.S.}$ stands for the Hilbert-Schmidt norm of a matrix M and $\nabla^2 f_h$ for the Hessian of f_h); see also [19]. Gaunt [12] later showed a generalization of this result, namely a version of (9) where the derivatives of order k of f_h can be bounded by derivatives of order $k-1$ of h. This still does not concur with the univariate case as we know that, in this case and when h' is bounded, one can bound one higher derivative of f_h: indeed, it holds $|f_h^{(3)}| \leqslant 2|h'|$ (here the function f_h is the solution (7) to the univariate version of the second order equation (6)). This loss of regularity is, however, not an artefact of the method of proof but is inherent to the method itself: Raič [18] exhibits a counterexample, namely a Lipschitz-continuous function such that the second derivative of f_h is *not* Lipschitz-continuous. We will discuss this example in detail later on.

1.2. **Multivariate Berry-Esseen bounds.** Let $(X_i)_{i\geqslant 1}$ be an i.i.d. sequence of random vectors in $\mathbb{R}^d$, and for simplicity take them centered with identity covariance matrix. Let $W = n^{-1/2}\sum_{i=1}^n X_i$, $Z \sim \mathcal{N}(0, I_d)$ and consider the problem of estimating $D(W, Z)$, some probability distance between the law of Z and that of W. According to [13], the earliest results on this problem in dimension $d \geq 2$ concern distances of the form (2) with $\mathcal{H}$ indicator functions of measurable convex sets in $\mathbb{R}^d$ (which is a multivariate generalization of the Kolmogorov distance). The best result for this choice of distance is due to [2] where an estimate of the form $d_{\mathcal{H}}(W, Z) \leq 400\, d^{1/4} n^{-1/2} E[|X_1|^3]$ is shown ($|\cdot|$ is the Euclidean norm); the dependence on the dimension is explicit and the best available for these moment assumptions and this distance. More recently, a high dimensional version of the same problem was studied in [9], with $\mathcal{H}$ the class of indicators of hyper-rectangles in $\mathbb{R}^d$; we also refer to the latter paper for an extensive and up-to-date literature review on such results.

Another important natural family of probability distances are the Wasserstein distances of order p (a.k.a. Mallows distances) defined as

$$\mathcal{W}_p(W, Z) = \left(\inf \mathbb{E}\left[|X_1 - Y_1|^p\right]\right)^{1/p} \tag{11}$$

where the infimum is taken over all joint distributions of the random vectors X_1 and Y_1 with respective marginals the laws of W and Z. Except in the case $p = 1$, such distances cannot be written under the form (2); as previously mentionned, when $p = 1$ the distance $\mathcal{W} := \mathcal{W}_1$ in (11) is of the form (8) with $\mathcal{H}$ the class of Lipschitz function with constant 1. Because $\mathcal{W}_p(\cdot,\cdot) \geq \mathcal{W}_{p'}(\cdot,\cdot)$ for all $p \geq p'$, bounds in p-Wasserstein distance are stronger than those in p'-Wasserstein distance; in particular $\mathcal{W}_p(\cdot,\cdot) \geq \mathcal{W}_1(\cdot,\cdot)$ for all $p \geq 1$. We refer to [24] for more information on p-Wasserstein distances. CLT's in Wasserstein distance have been studied, particularly in dimension 1, where we refer to the works [4, 20, 21] as well as [5] (and references therein) for convergence rates in p-Wasserstein for all $p \geq 1$ under the condition of existence of moments of order $2 + p$; in all cases the rate obtained is of optimal order $\mathcal{O}(1/\sqrt{n})$. In higher dimensions, results are also available in 2-Wasserstein distance, under more stringent assumptions on the X_i. For instance, Zhai [25] shows that when X_i is almost surely bounded, then a near-optimal rate of convergence in $\mathcal{O}(\log n/\sqrt{n})$ holds (this improves a result by Valiant et al. [23]). More recently, Courtade et al. [10] attained the optimal rate of convergence $\mathcal{O}(n^{-1/2})$, again in Wasserstein distance of order 2, under the assumption that X_i satisfies a Poincaré-type inequality; see also [11] for a similar result under assumption of log-concavity. Finally we mention the work of Bonis [6] where similar estimates are investigated (in Wasserstein-2 distance) under moment assumptions only; dependence of these estimates on the dimension is unclear (see [10, page 12]).

One of the key ingredients in many of the more recent above-mentioned references is the multivariate Stein's method. Rates of convergence in the multivariate CLT were first obtained Stein's method by Barbour in [1] (see also Götze [13]) whose methods (which rest on viewing the normal distribution as the stationary distribution of an Ornstein-Uhlenbeck diffusion, and using the generator of this diffusion as a characterizing operator) led to the so-called *generator approach to Stein's method* with starting point equation (6) and its solution given by the classical formula (7). Such an approach readily provides rates of convergence in *smooth-k*-Wasserstein distances, i.e. integral probability metrics of the form (2) with $\mathcal{H}$ $(= \mathcal{H}_{(k)})$ a set of smooth functions with derivatives up to some order k bounded by 1. Of course, the smaller the order k, the stronger the distance; in particular the case $k = 1$ coincides with the classical 1-Wasserstein distance (and therefore also (11) with $p = 1$). In [7, Theorem 3.1] it is proved that if X_i has a finite moment of order 4 then, for any smooth h,

$$\mathbb{E}[h(W) - h(Z)] \leqslant \frac{1}{\sqrt{n}}\left(\frac{1}{2}\sqrt{\mathbb{E}|X_i|^4 - d}\,\|\nabla h\|_\infty + \frac{\sqrt{2\pi}}{3}\mathbb{E}|X_i|^3 \sup_{x\in\mathbb{R}^d}\left\|\nabla^2 h(x)\right\|_{op}\right), \tag{12}$$

where $\|M\|_{op}$ denotes the operator norm of a matrix M and $|x|$ the Euclidean norm of a vector $x \in \mathbb{R}^d$. The rate $\mathcal{O}(1/\sqrt{n})$ is optimal. The fourth moment conditions are not optimal, nor is the restriction to twice differentiable test functions which implies that (12) does *not* lead to rates of convergence in the 1-Wasserstein distance. Similarly, the bounds on 2-Wasserstein distance recently obtained in [10, 11] are inspired by concepts related to Stein's method which were introduced in [14]; such an approach necessarily requires regularity assumptions on the density of the X_i. Hence no simple extension of their approach can lead to rates of convergence in Wasserstein distance with only moment conditions on the X_i (and in particular no smoothness assumptions on the densities). In other words, no optimal rates of convergence in Wasserstein distance are available under moment assumptions, and they seem out of reach if based on current available regularity results of Stein's equation.

1.3. **Contribution.** In this paper, we study Barbour's solution (7) to the Stein equation (6) and prove new regularity results: namely, if h is α-Hölder for some $0 < \alpha \leqslant 1$, then for all i, j, $\frac{\partial^2 f_h}{\partial x_i \partial x_j}$ is β-Hölder for $0 < \beta < \alpha$. Actually, we show the stronger estimate

$$\left|\frac{\partial^2 f_h}{\partial x_i \partial x_j}(x) - \frac{\partial^2 f_h}{\partial x_i \partial x_j}(y)\right| = \mathcal{O}\left(|x-y|^\alpha \log|x-y|\right), \tag{13}$$

for $|x-y|$ small. A precise statement, with explicit constants (which depend on α and on the dimension d), is given in Proposition 2.2. Note that from Shauder's theory, in the multivariate case (and contrary to the univariate one), one cannot hope in general for the second derivative of f_h to inherit the Lipschitz-regularity of h. Actually, Raič [18] gives a counter-example: if

$$h(x, y) = \max\{\min\{x, y\}, 0\},$$

then f_h defined by (7) is twice differentiable but $\frac{\partial^2 f_h}{\partial x \partial y}$ is not Lipschitz (whereas h is). We study this example in more detail in Proposition 2.4, which shows that (at least for $\alpha = 1$), the regularity (13) cannot be improved in general.

In a second step, we apply those regularity results to estimate the rate of convergence in the CLT, in Wasserstein distance.

Theorem 1.1. *Let $(X_i)_{i \geqslant 1}$ be an i.i.d. sequence of random vectors with unit covariance matrix, and $Z \sim \mathcal{N}(0, I_d)$. Assume that there exists $\delta \in (0,1)$ such that $\mathbb{E}[|X_i|^{2+\delta}] < \infty$. Then*

$$\mathcal{W}\left(n^{-1/2}\sum_{i=1}^n X_i, Z\right) \leqslant \frac{1}{n^{\frac{\delta}{2}}}\left[(K_1 + 2(1-\delta)^{-1})\mathbb{E}|X_i|^{2+\delta} + (K_2 + 2d(1-\delta)^{-1})\mathbb{E}|X_i|^{\delta}\right],$$

where $\mathcal{W}$ stands for the 1*-Wasserstein distance, and*

$$K_1 = 2^{3/2}\frac{2d+1}{d}\frac{\Gamma(\frac{1+d}{2})}{\Gamma(d/2)}$$
$$K_2 = 2\sqrt{\frac{2}{\pi}}\sqrt{d}.$$

Note that the rate in $\mathcal{O}(n^{-\delta/2})$ is optimal when only assuming moments of order $2+\delta$; see [3] or [16]. As mentioned previously, the bounds in Theorem 1.1 are to our knowledge the first optimal rates in 1-Wasserstein distance in the multidimensional case when assuming finite moments of order $2+\delta$ only.

From the previous Theorem is easily derived the following Corollary, which gives a near-optimal rate of order $\mathcal{O}(\log n/\sqrt{n})$ when X_i has finite moment of order 3.

Corollary 1.2. *Let $(X_i)_{i\geqslant 1}$ be an i.i.d. sequence of d-dimensional random vectors with unit covariance matrix. Assume that $\mathbb{E}[|X_i|^3] < \infty$. Then for $n \geqslant 3$,*

$$\mathcal{W}\left(n^{-1/2}\sum_{i=1}^n X_i, Z\right) \leqslant e\,\frac{C(d) + 2(1+d)\,\log n}{\sqrt{n}}\,\mathbb{E}|X_i|^3,$$

where $C(d) = 2^{3/2}\frac{2d+1}{d}\frac{\Gamma(\frac{1+d}{2})}{\Gamma(d/2)} + 2\sqrt{\frac{2}{\pi}}\sqrt{d}$.

Compared to [25], our assumption on the distribution of X_i is much weaker; however the distance used in [25] is stronger and the constants are sharper ([25] obtains a constant in $\mathcal{O}(\sqrt{d})$). [10] has the advantage of stronger rate of convergence (it is optimal when ours is near-optimal) and stronger distance, but the drawback of a less tractable assumption on the distribution of X_i (it should satisfy a Poincaré or weighted Poincaré inequality).

2. Regularity of solutions of Stein's equation

Throughout the rest of the paper, for $x, y \in \mathbb{R}^d$, we denote by $x \cdot y$ the Euclidean scalar product between x and y, and $|x|$ the Euclidean norm of x. For a matrix M of size $d \times d$, its operator norm is defined as

$$\|M\|_{op} = \sup_{x\in\mathbb{R}^d;\,|x|=1} |Mx|.$$

Define the α-Hölder semi-norm, for $\alpha \in (0,1]$, by

$$[h]_\alpha = \sup_{x\neq y}\frac{|h(x)-h(x)|}{|x-y|^\alpha}.$$

For a multi-index $\mathbf{i} = (i_1,\dots,i_d) \in \mathbb{N}^d$, the multivariate Hermite polynomial $H_\mathbf{i}$ is defined by

$$H_\mathbf{i}(x) = (-1)^{|\mathbf{i}|}e^{|x|^2/2}\frac{\partial^{|\mathbf{i}|}}{\partial x_1^{i_1}\dots\partial x_d^{i_d}}e^{-|x|^2/2},$$

where $|\mathbf{i}| = i_1 + \ldots + i_d$.

Let $h : \mathbb{R}^d \to \mathbb{R}$, and f_h be defined by (when the integral makes sense)

$$f_h(x) = -\int_0^1 \frac{1}{2t}\mathbb{E}\,\bar{h}(Z_{x,t})\,dt, \tag{14}$$

where

$$\bar{h}(x) = h(x) - \mathbb{E}\,h(Z),$$

and

$$Z_{x,t} = \sqrt{t}\,x + \sqrt{1-t}\,Z.$$

Recall that, when h is smooth with compact support, then (14) defines a solution to the Stein equation (6), see [1, 7]. We shall prove that this is still the case when only assuming Hölder-regularity of h.

Proposition 2.1. *Let $h : \mathbb{R}^d \to \mathbb{R}$ be a α-Hölder function; that is, $[h]_\alpha < \infty$. Let f_h be the function given by* (14)*. Then:*

- *f_h is twice differentiable and for $\mathbf{i} = (i_1,\dots,i_d) \in \mathbb{N}^d$ such that $1 \leqslant |\mathbf{i}| \leqslant 2$,*

$$\frac{\partial^{|\mathbf{i}|}f_h}{\partial x_1^{i_1}\dots\partial x_d^{i_d}} = -\int_0^1 \frac{t^{\frac{|\mathbf{i}|}{2}-1}}{2(1-t)^{\frac{|\mathbf{i}|}{2}}}\mathbb{E}[H_\mathbf{i}(Z)\bar{h}(Z_{x,t})]dt. \tag{15}$$

- *f_h is a solution to the Stein equation* (6)*.*

Proof. Fix $t \in (0,1)$. Recall $\omega(x) = (2\pi)^{-d/2} e^{-|x|^2/2}$ is the density of the standard d-dimensional gaussian measure. Since

$$\mathbb{E}\,\bar{h}(Z_{x,t}) = \int_{\mathbb{R}^d} \bar{h}(\sqrt{t}\,x + \sqrt{1-t}\,z)\,\omega(z)\,dz = (-1)^d t^{d/2}(1-t)^{-d/2} \int_{\mathbb{R}^d} \bar{h}(u)\,\omega\left(\frac{u-\sqrt{t}\,x}{\sqrt{1-t}}\right)\,du,$$

we have, from Lebesgue's derivation theorem, and another change of variable,

$$\frac{\partial^{|\mathbf{i}|}}{\partial x_1^{i_1}\dots\partial x_d^{i_d}}\mathbb{E}\,\bar{h}(Z_{x,t}) = \frac{t^{\frac{|\mathbf{i}|}{2}}}{(1-t)^{\frac{|\mathbf{i}|}{2}}}\mathbb{E}[H_{\mathbf{i}}(Z)\bar{h}(Z_{x,t})]. \tag{16}$$

Now note that by α-Hölder regularity, and using the fact that $\mathbb{E}H_{\mathbf{i}}(Z_1,\dots,Z_d) = 0$,

$$\begin{aligned} |\mathbb{E}H_{\mathbf{i}}(Z)\bar{h}(Z_{x,t})| &= |\mathbb{E}H_{\mathbf{i}}(Z)(\bar{h}(Z_{x,t}) - \bar{h}(\sqrt{t}\,x))| \\ &\leqslant \mathbb{E}\left[|H_{\mathbf{i}}(Z)|\,|Z|^\alpha\right](1-t)^{\alpha/2}. \end{aligned} \tag{17}$$

Thus we can apply again Lebesgue's derivation theorem to obtain (15).

Now let $\omega_t(x) = t^{-d/2}\omega\left(\frac{x}{\sqrt{t}}\right)$; ω_{1-t} is the density of $\sqrt{1-t}\,Z$. It is well known (and can be easily checked) that ω_t solves the heat equation

$$\partial_t\,\omega_t = \frac{1}{2}\,\Delta\,\omega_t.$$

We deduce (again applying Lebesgue's derivation theorem, valid since $\bar{h}$ has polynomial growth at infinity) that

$$\begin{aligned} \partial_t\mathbb{E}\,\bar{h}(Z_{x,t}) &= \partial_t \int_{\mathbb{R}^d} \bar{h}(u)\,\omega_{1-t}(u-\sqrt{t}\,x)\,du \\ &= -\int_{\mathbb{R}^d} \bar{h}(u)\,\partial_t\omega_{1-t}(u-\sqrt{t}\,x)\,du - \frac{1}{2\sqrt{t}}\int_{\mathbb{R}^d} \bar{h}(u)\nabla\omega_{1-t}(u-\sqrt{t}\,x)\cdot x\,du \\ &= -\frac{1}{2}\int_{\mathbb{R}^d} \bar{h}(u)\Delta\,\omega_{1-t}(u-\sqrt{t}\,x)\,du - \frac{1}{2\sqrt{t}}\int_{\mathbb{R}^d} \bar{h}(u)\nabla\omega_{1-t}(u-\sqrt{t}\,x)\cdot x\,du \\ &= -\frac{1}{2t}\Delta_x \int_{\mathbb{R}^d} \bar{h}(u)\,\omega_{1-t}(u-\sqrt{t}\,x)\,du + \frac{1}{2t}\nabla_x\left[\int_{\mathbb{R}^d} \bar{h}(u)\,\omega_{1-t}(u-\sqrt{t}\,x)\,du\right]\cdot x \\ &= -\frac{1}{2t}(\Delta - x\cdot\nabla)\mathbb{E}\,\bar{h}(Z_{x,t}). \end{aligned}$$

Finally,

$$\bar{h}(x) = \int_0^1 \partial_t\mathbb{E}\,\bar{h}(Z_{x,t})dt = -\int_0^1 \frac{1}{2t}(\Delta - x\cdot\nabla)\mathbb{E}\,\bar{h}(Z_{x,t})\,dt = (\Delta - x\cdot\nabla)f_h,$$

the last equality being justified by (16) and the bound (17). □

Before stating our main regularity results, let us give the idea behind the proof. Starting from (15), we have that

$$\frac{\partial^2 f_h}{\partial x_i\partial x_j}(x) - \frac{\partial^2 f_h}{\partial x_i\partial x_j}(y) = -\int_0^1 \frac{1}{2(1-t)}\mathbb{E}[(Z_iZ_j - \delta_{ij})(\bar{h}(Z_{x,t}) - \bar{h}(Z_{y,t}))]dt.$$

Using the α-Hölder regularity of h, the modulus of the integrand in last integral can be bounded by

$$\frac{1}{2(1-t)}\mathbb{E}[|Z_iZ_j - \delta_{ij}|\,|\bar{h}(Z_{x,t}) - \bar{h}(Z_{y,t})|] \leqslant C_{ij}\frac{t^{\alpha/2}}{1-t}|x-y|^\alpha,$$

C_{ij} being some constant. However, the function in the right hand-side is not integrable for t close to 1. Thus, for $\eta > 0$, we split the integral between 0 and $1-\eta$ on the one hand (where we can use lour previous bound), and between $1-\eta$ and 1 on the other hand. To bound the second integral, we remark that since $\mathbb{E}[Z_iZ_j-\delta_{ij}]=0$, we have that

$$\mathbb{E}[(Z_iZ_j-\delta_{ij})\bar{h}(Z_{x,t})] = \mathbb{E}[(Z_iZ_j-\delta_{ij})(\bar{h}(Z_{x,t})-\bar{h}(\sqrt{t}x))],$$

which, in modulus, is less than (again using the regularity of h)

$$\mathbb{E}[|Z_iZ_j-\delta_{ij}|\,\|Z\|^\alpha](1-t)^{\alpha/2}.$$

The power $(1-t)^{\alpha/2}$ that is gained makes the integral converge. Finally, we optimize in $\eta>0$.

We are concerned, however, in obtaining the best constants possible (seen as functions of the dimension d and α); this tends to make the proofs more technical that needed if one is only concerned with showing regularity. For this reason, the detailed exposition of the proof in full detail is deferred to Section 5.

We start with the regularity in terms of the operator norm of the Hessian of f_h.

Proposition 2.2. *Let $h:\mathbb{R}^d\to\mathbb{R}$ be a α-Hölder function for some $\alpha\in(0,1]$. Then the solution f_h (7) of the Stein equation (6) satisfies:*

$$\begin{aligned}\|\nabla^2 f_{|x}-\nabla^2 f_{|y}\|_{op} &\leqslant [h]_\alpha|x-y|^\alpha\left(C_1(\alpha,d)-2\log|x-y|\right), && \text{if } |x-y|\leqslant 1\\ &\leqslant C_1(\alpha,d)\,[h]_\alpha && \text{if } |x-y|>1,\end{aligned} \tag{18}$$

where

$$C_1(\alpha,d)=2^{\frac{\alpha}{2}+1}\frac{\alpha+2d}{\alpha d}\frac{\Gamma(\frac{\alpha+d}{2})}{\Gamma(d/2)}. \tag{19}$$

In particular, for all $0<\beta<\alpha$, $\frac{\partial^2 f_h}{\partial x_i\partial x_j}$ is globally β-Hölder:

$$\|\nabla^2 f_{|x}-\nabla^2 f_{|y}\|_{op}\leqslant\left(C_1(\alpha,d)+\frac{2}{\alpha-\beta}\right)|x-y|^\beta[h]_\alpha. \tag{20}$$

It also holds the $(1+\log)$ α-Hölder regularity

$$\|\nabla^2 f_{|x}-\nabla^2 f_{|y}\|_{op}\leqslant|x-y|^\alpha\left(C_1(\alpha,d)+|\log|x-y||\right)[h]_\alpha. \tag{21}$$

Now we turn to the regularity of the Laplacian.

Proposition 2.3. *Let $h:\mathbb{R}^d\to\mathbb{R}$ be a α-Hölder function for some $\alpha\in(0,1]$. Then the solution f_h (7) of the Stein equation (6) satisfies:*

$$\begin{aligned}\left|\Delta f_{|x}-\Delta f_{|y}\right| &\leqslant [h]_\alpha|x-y|^\alpha\left(C_2(\alpha,d)-2\,d\,\log|x-y|\right), && \text{if } |x-y|\leqslant 1\\ &\leqslant C_2(\alpha,d)\,[h]_\alpha && \text{if } |x-y|>1,\end{aligned}$$

where

$$\begin{aligned}C_2(\alpha,d) &= 2^{\frac{\alpha}{2}+1}\tfrac{(\alpha+2d)\,\Gamma(\frac{\alpha+d}{2})}{\alpha\,\Gamma(d/2)} \text{ if } \alpha\in(0,1),\\ C_2(1,d) &= 2\sqrt{\tfrac{2}{\pi}}\sqrt{d}.\end{aligned} \tag{22}$$

In particular, for all $0<\beta<\alpha$,

$$|\Delta f_{|x}-\Delta f_{|y}|\leqslant\left(C_2(\alpha,d)+d\frac{2}{\alpha-\beta}\right)|x-y|^\beta[h]_\alpha. \tag{23}$$

Note that Proposition 2.2 implies that, for $\alpha=1$, when $|x-y|$ is small, then

$$\left|\frac{\partial^2 f_h}{\partial x_i\partial x_j}(x)-\frac{\partial^2 f_h}{\partial x_i\partial x_j}(y)\right|=\mathcal{O}\left(|x-y|\log|x-y|\right).$$

The example given by Raič [18] shows that this rate is optimal; indeed, we have the following result.

Proposition 2.4. *Let $h : \mathbb{R}^2 \to \mathbb{R}$ be the Lipschitz function defined by $h(x,y) = \max(0, \min(x,y))$. Then*

$$\frac{\partial^2 f_h}{\partial x \partial y}(u,u) - \frac{\partial^2 f_h}{\partial x \partial y}(0,0) \underset{u\to 0^+}{\sim} \frac{1}{\sqrt{2\pi}}\, u \log u.$$

The proof can be found in the Appendix.

3. Multivariate Berry-Esseen bounds in Wasserstein distance

As anticipated, we apply the regularity results obtained in previous section to obtain Berry-Esseen bounds in the CLT, in 1-Wasserstein distance.

Let $X_1, X_2, \ldots$ be an i.i.d. sequence of centered, square-integrable and isotropic random vectors; that is, $\mathbb{E}[X_1] = 0$ and $\mathbb{E}[X_1 X_1^T] = I_d$. Let $W = n^{-1/2}\sum_{i=1}^n X_i$. We are interested in $\mathcal{W}_\alpha(W,Z)$ for $\alpha \in (0,1]$, where the α-Wasserstein distance is defined as

$$\mathcal{W}_\alpha(X,Y) = \sup_{\{h \in C(\mathbb{R}^n, \mathbb{R}) | [h]_\alpha \leqslant 1\}} \mathbb{E}h(X) - \mathbb{E}h(Y).$$

As in the introduction, for $\alpha = 1$, the resulting distance is $\mathcal{W} := \mathcal{W}_1$, the classical 1-Wasserstein distance (that is, $\mathcal{W}(X,Y) = \sup_{\mathcal{H}} \mathbb{E}h(X) - \mathbb{E}h(Y)$ with $\mathcal{H}$ the collection of 1-Lipschitz functions).

We are now in a position to prove our main Theorem. We first give a more general version of it in α-Wasserstein distances; Theorem 1.1 is just the following Theorem applied to $\alpha = 1$.

Theorem 3.1. *Let $\alpha \in (0,1]$ and $(X_i)_{i\geqslant 1}$ be an i.i.d. sequence of d-dimensional random vectors with unit covariance matrix. Assume that there exists $\delta \in (0,\alpha)$ such that $\mathbb{E}[|X_i|^{2+\delta}] < \infty$. Then*

$$\mathcal{W}_\alpha\left(n^{-1/2}\sum_{i=1}^n X_i, Z\right) \leqslant \frac{1}{n^{\frac{\delta}{2}}}\left[\left(C_1(\alpha,d) + \frac{2}{\alpha-\delta}\right)\mathbb{E}|X_i|^{2+\delta} + \left(C_2(\alpha,d) + d\frac{2}{\alpha-\delta}\right)\mathbb{E}|X_i|^{\delta}\right],$$

where $C_1(\alpha,d)$ and $C_2(\alpha,d)$ are respectively defined in (19) *and* (22).

Remark 3.2. *From Stirling's formula, $C_1(\alpha,d) = \mathcal{O}(\sqrt{d})$, $C_2(\alpha,d) = \mathcal{O}(d^{1+\alpha/2})$ for $\alpha \in (0,1)$ and $C_2(1,d) = \mathcal{O}(\sqrt{d})$.*

Proof of Theorems 1.1 and 3.1 . Let h be α-Hölder (with $[h]_\alpha \leqslant 1$) and f_h be the solution of the Stein equation defined by Proposition 2.1. Then,

$$\begin{aligned}\mathbb{E}[h(W) - h(Z)] &= \mathbb{E}[\Delta f_h(W) - W\cdot\nabla f_h(W)] \\ &= \frac{1}{n}\sum_{i=1}^n \left[\mathbb{E}[\Delta f_h(W) - \sqrt{n}X_i\cdot\nabla f_h(W)\right].\end{aligned}$$

The following calculations already appeared in the literature (see e.g. [18]), we include them here for completeness. Let $W_i = W - X_i/\sqrt{n} = \frac{1}{\sqrt{n}}\sum_{j\neq i} X_j$. By Taylor's formula, we have for some uniformly distributed in $[0,1]$ (and independent of everything else) θ

$$\mathbb{E}[X_i \cdot \nabla f_h(W)] = \frac{1}{\sqrt{n}}\mathbb{E}\left[X_i^T \nabla^2 f_h\left(W_i + \theta\frac{X_i}{\sqrt{n}}\right)X_i\right],$$

leading to

$$\mathbb{E}[h(W) - h(Z)] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\Delta f_h(W) - X_i^T\nabla^2 f_h\left(W_i + \theta\frac{X_i}{\sqrt{n}}\right)X_i\right].$$

Let $X_{i,j}$ be the jth coordinate of X_i. Since W_i is independent of X_i, and X_i has unit covariance matrix, we have

$$\mathbb{E}\left[X_i^T\nabla^2 f_h(W_i)X_i\right] = \sum_{j,k=1}^d \mathbb{E}\left[X_{i,j}X_{i,k}\frac{\partial^2 f_h}{\partial x_j\partial x_k}(W_i)\right] = \sum_{j=1}^d \mathbb{E}\left[\frac{\partial^2 f_h}{\partial x_j^2}(W_i)\right] = \mathbb{E}[\Delta f_h(W_i)].$$

Finally,

$$\mathbb{E}[h(W)-h(Z)] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\Delta f_h(W)-\Delta f_h(W_i) - X_i^T\left(\nabla^2 f_h\left(W_i+\theta\frac{X_i}{\sqrt{n}}\right)-\nabla^2 f_h(W_i)\right)X_i\right].$$

Note then that

$$\left|X_i^T\left(\nabla^2 f_h\left(W_i+\theta\frac{X_i}{\sqrt{n}}\right)-\nabla^2 f_h(W_i)\right)X_i\right| \leqslant |X_i|^2\left\|\nabla^2 f_h\left(W_i+\theta\frac{X_i}{\sqrt{n}}\right)-\nabla^2 f_h(W_i)\right\|_{op}.$$

We now use Propositions 2.2 and 2.3 with $\beta=\delta\leqslant\alpha$ to obtain:

$$\mathbb{E}[h(W)-h(Z)] \leqslant \frac{1}{n}\sum_{i=1}^{n}\left[\left(C_2(\alpha,d)+d\frac{2}{\alpha-\delta}\right)\frac{\mathbb{E}[|X_i|^{\delta}\theta^{\delta}]}{n^{\delta/2}}+\left(C_1(\alpha,d)+\frac{2}{\alpha-\delta}\right)\frac{\mathbb{E}[|X_i|^{2+\delta}\theta^{\delta}]}{n^{\delta/2}}\right].$$

Noting that $\mathbb{E}[\theta^{\beta}]\leqslant 1$ and rearranging, we obtain the result. □

Corollary 3.3. *Let $\alpha\in(0,1]$, and $(X_i)_{i\geqslant 1}$ be an i.i.d. sequence of d-dimensional random vectors with unit covariance matrix. Assume that $\mathbb{E}[|X_i|^{2+\alpha}]<\infty$. Then for $n>\exp(2/\alpha)$,*

$$\mathcal{W}_{\alpha}\left(n^{-1/2}\sum_{i=1}^{n}X_i, Z\right)\leqslant e\,\frac{C_1(\alpha,d)+C_2(\alpha,d)+2(1+d)\,\log n}{n^{\frac{\alpha}{2}}}\,\mathbb{E}|X_i|^{2+\alpha},$$

where $C_1(\alpha,d)$ and $C_2(\alpha,d)$ are respectively defined in (19) and (22).

Proof. By Hölder's and the Cauchy-Schwarz inequalities, for any $\delta\leqslant\alpha$, $\mathbb{E}|X_i|^{\delta}\leqslant\left(\mathbb{E}|X_i|^{2+\alpha}\right)^{\delta/(2+\alpha)}$. But by Jensen's inequality, $\mathbb{E}|X_i|^{2+\alpha}\geqslant(\mathbb{E}|X_i|^2)^{(2+\alpha)/2}=d^{(2+\alpha)/2}\geqslant 1$, so that, since $\delta/(2+\alpha)<1$, $\left(\mathbb{E}|X_i|^{2+\alpha}\right)^{\delta/(2+\alpha)}\leqslant\mathbb{E}|X_i|^{2+\alpha}$. Similarly, $\mathbb{E}|X_i|^{2+\delta}\leqslant\left(\mathbb{E}|X_i|^{2+\alpha}\right)^{1-\frac{\alpha-\delta}{2+\alpha}}\leqslant\mathbb{E}|X_i|^{2+\alpha}$. Note now that the bound of Theorem 3.1 holds for any $0<\delta<\alpha$. Choosing $\alpha-\delta=2/\log n$ achieves the proof since $n^{-\frac{1}{\log n}}=1/e$. □

When applied to $\alpha=1$, previous corollary leads to Corollary 1.2, which we recall here: as long as $\mathbb{E}|X_i|^3<\infty$,

$$\mathcal{W}\left(n^{-1/2}\sum_{i=1}^{n}X_i, Z\right)\leqslant e\,\frac{C(d)+2(1+d)\,\log n}{\sqrt{n}}\,\mathbb{E}|X_i|^3,$$

where $C(d)=2^{3/2}\frac{(2d+1)\,\Gamma(\frac{d+1}{2})}{d\Gamma(\frac{d}{2})}+2\sqrt{\frac{2}{\pi}}\sqrt{d}$. [25] also obtains a near-optimal rate of convergence in $\mathcal{O}(\log n/\sqrt{n})$, but under the much stronger assumption that $|X_i|\leqslant\beta$ almost surely; nevertheless, the distance used in [25] (the quadratic Wasserstein distance) is stronger than ours, the behaviour of the constant on the higher order term is $\mathcal{O}(\sqrt{d})$, here we obtain $\mathcal{O}(d)$.

4. Extension to higher order derivatives

The regularity result easily extends to higher order derivatives.

Proposition 4.1. *Let $h:\mathbb{R}^d\to\mathbb{R}$ be a smooth, compactly supported function, and denote by $[h]_{\alpha,p}$ a common α-Hölder constant for all derivatives of order p of h. Then the solution f_h (7) of the Stein equation (6) satisfies, for all $(i_1,\dots,i_{p+2})\in\{1,\dots,d\}^{p+2}$:*

$$\begin{aligned}\left|\frac{\partial^{p+2}f}{\prod_{j=1}^{p+2}\partial x_{i_j}}(x)-\frac{\partial^{p+2}f}{\prod_{j=1}^{p+2}\partial x_{i_j}}(y)\right| &\leqslant [h]_{\alpha,p}|x-y|^{\alpha}\left(A-2\log|x-y|\right), &&\text{if } |x-y|\leqslant 1\\ &\leqslant A\,[h]_{\alpha,p} &&\text{if } |x-y|>1,\end{aligned}$$

where

$$A=2^{\alpha/2+1}\frac{\alpha+d+1}{\alpha}\frac{\Gamma(\frac{\alpha+d}{2})}{\Gamma(d/2)}.$$

In particular, all derivatives of the order $p+2$ of f_h are β-Hölder for any $0<\beta<\alpha$ and we have

$$\left|\frac{\partial^{p+2} f}{\prod_{j=1}^{p+2}\partial x_{i_j}}(x) - \frac{\partial^{p+2} f}{\prod_{j=1}^{p+2}\partial x_{i_j}}(y)\right| \leqslant [h]_{\alpha,p}\left(A + \frac{2}{\alpha-\beta}\right)|x-y|^{\beta}.$$

Proof. Taking derivatives in (7), we have

$$\frac{\partial^{p+2} f}{\prod_{j=1}^{p+2}\partial x_{i_j}}(x) = \int_0^1 \frac{t^p}{2}\mathbb{E}\left[\frac{\partial^{p+2}\bar{h}}{\prod_{j=1}^{p+2}\partial x_{i_j}}(Z_{x,t})\right]dt.$$

Next perform two Gaussian integration by parts against two indices i_{p+1} and i_{p+2}, say, to get

$$\frac{\partial^{p+2} f}{\prod_{j=1}^{p+2}\partial x_{i_j}}(x) = \int_0^1 \frac{t^p}{2(1-t)}\mathbb{E}\left[(Z_{i_{p+1}}Z_{i_{p+2}} - \delta_{i_{p+1}i_{p+2}})\frac{\partial^{p}\bar{h}}{\prod_{j=1}^{p}\partial x_{i_j}}(Z_{x,t})\right]dt.$$

Then, using the same method as in the proof of Proposition 2.2 (we do not give all the details here), we have

$$\left|\frac{\partial^{p+2} f}{\prod_{j=1}^{p+2}\partial x_{i_j}}(x) - \frac{\partial^{p+2} f}{\prod_{j=1}^{p+2}\partial x_{i_j}}(y)\right| \leqslant -[h]_{\alpha,p}\log\eta + [h]_{\alpha,p}2^{\alpha/2+1}\frac{\alpha+d+1}{\alpha}\frac{\Gamma(\frac{\alpha+d}{2})}{\Gamma(d/2)}\eta^{\alpha/2}.$$

Choose $\eta = |x-y|$ if $|x-y| \leqslant 1$, 1 otherwise, to get the first result, and the fact that $-\log u \leqslant \frac{1}{\alpha-\beta}u^{\beta-\alpha}$ if $u \leqslant 1$ for the second one. □

We stress that one possible application of this Proposition would be a multivariate Berry-Esseen bound in the CLT with matching moments (i.e. assuming that the underlying random variables X_i share the same first k moments with the Gaussian). In this case, faster rates of convergence are expected, see [12].

5. The remaining proofs

Proof of Proposition 2.2. Recall that

$$\frac{\partial^2 f_h}{\partial x_i \partial x_j} = -\int_0^1 \frac{1}{2(1-t)}\mathbb{E}[(Z_iZ_j - \delta_{ij})\bar{h}(Z_{x,t})]dt.$$

Since $\mathbb{E}[Z_iZ_j - \delta_{ij}] = 0$, we have $\mathbb{E}[(Z_iZ_j - \delta_{ij})\bar{h}(\sqrt{t}x)] = 0$, so that

$$\mathbb{E}[(Z_iZ_j - \delta_{ij})\bar{h}(Z_{x,t})] = \mathbb{E}[(Z_iZ_j - \delta_{ij})(\bar{h}(Z_{x,t}) - \bar{h}(\sqrt{t}x))]. \tag{24}$$

Thus,

$$\nabla^2 f_h(x) = -\int_0^1 \frac{1}{2(1-t)}\mathbb{E}[(ZZ^T - I_d)(\bar{h}(Z_{x,t}) - \bar{h}(\sqrt{t}x))]\,dt,$$

where Z^T denotes the transpose of Z. Let $a = (a_1,\dots,a_d)^T \in \mathbb{R}^d$ and assume that $|a| = 1$. We have

$$\begin{aligned} a^T\nabla^2 f_h(x)a &= -\int_0^1 \frac{1}{2(1-t)}\mathbb{E}[a^T(ZZ^T - I_d)\,a\,(\bar{h}(Z_{x,t}) - \bar{h}(\sqrt{t}x))]dt \\ &= -\int_0^1 \frac{1}{2(1-t)}\mathbb{E}[((Z\cdot a)^2 - 1)\,(\bar{h}(Z_{x,t}) - \bar{h}(\sqrt{t}x))]dt. \end{aligned}$$

Since $|\bar{h}(Z_{x,t}) - \bar{h}(\sqrt{t}x)| \leqslant [h]_\alpha(1-t)^{\alpha/2}\|Z\|^\alpha$, we also have

$$|\mathbb{E}[((Z\cdot a)^2 - 1)\,(\bar{h}(Z_{x,t}) - \bar{h}(\sqrt{t}x))]\,| \leqslant [h]_\alpha\mathbb{E}[|(a\cdot Z)^2 - 1|\,\|Z\|^\alpha](1-t)^{\alpha/2}. \tag{25}$$

Let us bound $E[|(a\cdot Z)^2-1|\,\|Z\|^\alpha]$. Let $(a,e_2,\dots,e_d)$ be an orthonormal basis and $\tilde Z=(a\cdot Z,e_2\cdot Z,\dots,e_d\cdot Z)^T$. Then $\tilde Z\sim\mathcal N(0,I_d)$. Moreover, $E[|(a\cdot Z)^2-1|\,\|Z\|^\alpha]=\mathbb E[|\tilde Z_1^2-1|\|\tilde Z\|^\alpha]$. Thus,

$$\begin{aligned}E[|(a\cdot Z)^2-1|\,\|Z\|^\alpha]&=\mathbb E[|\tilde Z_1^2-1|\|\tilde Z\|^\alpha]\\&\leqslant\mathbb E[(\tilde Z_1^2+1)\|\tilde Z\|^\alpha]\\&=\frac1d\sum_{i=1}^d\mathbb E[(\tilde Z_i^2+1)\|\tilde Z\|^\alpha]\\&=\frac1d\mathbb E[(\|\tilde Z\|^2+d)\|\tilde Z\|^\alpha].\end{aligned}$$

For all $\beta>0$, $\mathbb E\|Z\|^\beta=\frac{2^{\frac\beta2}\Gamma(\frac{\beta+d}{2})}{\Gamma(d/2)}$. We define

$$C=\frac1d\mathbb E[(\|\tilde Z\|^2+d)\,\|\tilde Z\|^\alpha]=2^{\frac\alpha2}\frac{\alpha+2d}{d}\frac{\Gamma(\frac{\alpha+d}{2})}{\Gamma(d/2)}. \tag{26}$$

This shows in particular that $\|\nabla^2 f_h(x)\|_{op}$ is bounded.

Now we consider $\left|a^T\left(\nabla^2 f_h(x)-\nabla^2 f_h(y)\right)a\right|$ and split the integral into two parts. Let $\eta\in[0,1]$. We have

$$\begin{aligned}&|a^T\left(\nabla^2 f_h(x)-\nabla^2 f_h(y)\right)a|\\&=\left|\int_0^1\frac{1}{2(1-t)}\mathbb E\left[a^T(ZZ^T-I_d)\,a(\bar h(Z_{x,t})-\bar h(Z_{y,t})\right]dt\right|\\&\leqslant\int_0^{1-\eta}\frac{1}{2(1-t)}\mathbb E\left[|a^T(ZZ^T-I_d)\,a|\,|\bar h(Z_{x,t})-\bar h(Z_{y,t})|\right]dt\\&+\left|\int_{1-\eta}^1\frac{1}{2(1-t)}\mathbb E\left[a^T(ZZ^T-I_d)\,a(\bar h(Z_{x,t})-\bar h(Z_{y,t}))\right]\right|dt.\end{aligned}$$

Using the α-Hölder regularity of h for the first part of the integral and (24) twice in the second part together with (25) and (26), we can bound the previous quantity by

$$[h]_\alpha|x-y|^\alpha\mathbb E\left[|(a\cdot Z)^2-1|\right]\int_0^{1-\eta}\frac{t^{\alpha/2}}{2(1-t)}dt+[h]_\alpha C\int_{1-\eta}^1(1-t)^{-1+\alpha/2}dt \tag{27}$$

$$\leqslant[h]_\alpha\left(-|x-y|^\alpha\log\eta+\frac{2C}{\alpha}\eta^{\alpha/2}\right), \tag{28}$$

where to obtain (28), we used the facts that $\mathbb E\left[|(a\cdot Z)^2-1|\right]\leqslant 2$ and $t^{\alpha/2}\leqslant 1$. Choose $\eta=|x-y|^2$ if $|x-y|\leqslant 1$, $\eta=1$ otherwise to get (18). Equation (21) is a straightforward reformulation since $1+|\log(u)|\geqslant 1$. To get (20), simply note that for $0<\beta<\alpha$ and $0<u\leqslant 1$, $-\log u\leqslant\frac{1}{\alpha-\beta}u^{\beta-\alpha}$ and for $1\leqslant u$, $1\leqslant u^\beta$. $\square$

Proof of Proposition 2.3. The regularity of the Laplacian is proved in a similar manner as for the operator norm of the Hessian; we do not detail the computations here. Let $\alpha \in (0,1)$. We have

$$\begin{aligned}
&|\Delta f_{|x} - \Delta f_{|y}| \\
&= \left| \int_0^1 \frac{1}{2(1-t)} \mathbb{E}\left[\sum_{i=1}^d (Z_i^2 - 1)(\bar{h}(Z_{x,t}) - \bar{h}(Z_{y,t}) \right] dt \right| \\
\leqslant & \int_0^{1-\eta} \frac{1}{2(1-t)} \mathbb{E}\left[\left| \sum_{i=1}^d (Z_i^2 - 1) \right| |\bar{h}(Z_{x,t}) - \bar{h}(Z_{y,t})| \right] dt \\
&+ \int_{1-\eta}^1 \left| \frac{1}{2(1-t)} \mathbb{E}\left[\sum_{i=1}^d (Z_i^2 - 1)(\bar{h}(Z_{x,t}) - \bar{h}(Z_{y,t})) \right] \right| dt \\
\leqslant & [h]_\alpha |x-y|^\alpha \mathbb{E}\left[\|Z\|^2 + d \right] \int_0^{1-\eta} \frac{t^{\alpha/2}}{2(1-t)} dt + [h]_\alpha \mathbb{E}[(\|Z\|^2 + d)\|Z\|^\alpha] \int_{1-\eta}^1 (1-t)^{-1+\alpha/2} dt \\
\leqslant & [h]_\alpha \left(-d\,|x-y|^\alpha \log \eta + \frac{2\mathbb{E}[(\|Z\|^2 + d)\|Z\|^\alpha]}{\alpha} \eta^{\alpha/2} \right).
\end{aligned}$$

Note that

$$\mathbb{E}[(\|Z\|^2 + d)\|Z\|^\alpha] = \frac{2^{\frac{\alpha}{2}+1}\Gamma(\frac{\alpha+d}{2}+1) + d\, 2^{\frac{\alpha}{2}}\Gamma(\frac{\alpha+d}{2})}{\Gamma(d/2)} = 2^{\frac{\alpha}{2}}(\alpha + 2d)\frac{\Gamma(\frac{\alpha+d}{2})}{\Gamma(d/2)},$$

and choose again $\eta = |x-y|^2$ if $|x-y| \leqslant 1$, $\eta = 1$ otherwise.

We can obtain better constants in the case $\alpha = 1$. Indeed, note that by using only one integration by parts,

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^d (Z_i^2 - 1)(\bar{h}(Z_{x,t}) - \bar{h}(Z_{y,t})) \right] &= \sqrt{1-t}\,\mathbb{E}\left[\sum_{i=1}^d Z_i(\partial_i \bar{h}(Z_{x,t}) - \partial_i \bar{h}(Z_{y,t})) \right] \\
&= \sqrt{1-t}\,\mathbb{E}\left[Z \cdot (\nabla h(Z_{t,x}) - \nabla h(Z_{t,y}) \right],
\end{aligned}$$

whose modulus can be thus bounded by

$$2\sqrt{1-t}\,\mathbb{E}[\|Z\|] = \sqrt{1-t}\frac{2\sqrt{2}}{\sqrt{\pi}}\sqrt{d}.$$

Using this bound in the integral between $1-\eta$ and 1, and choosing η as in Proposition 2.2, we obtain the results. □

Proof of Proposition 2.4. Let $u > 0$. Denote $Z_i^{t,u} = \sqrt{t}\,u + \sqrt{1-t}\,Z_i$. We have

$$\begin{aligned}
\frac{\partial^2 f_h}{\partial x \partial y}(u,u) &= -\int_0^1 \frac{1}{2(1-t)} \mathbb{E}[Z_1 Z_2 h(Z_1^{t,u}, Z_2^{t,u})]\, dt \\
&= -\int_0^1 \frac{1}{2(1-t)} \mathbb{E}\left[Z_1 Z_2 (\mathbf{1}_{Z_2^{t,u} \geqslant Z_1^{t,u} \geqslant 0} Z_1^{t,u} + \mathbf{1}_{Z_1^{t,u} \geqslant Z_2^{t,u} \geqslant 0} Z_2^{t,u}) \right] dt \\
&= -\int_0^1 \frac{1}{1-t} \mathbb{E}\left[Z_1 Z_2 \mathbf{1}_{Z_2^{t,u} \geqslant Z_1^{t,u} \geqslant 0}\, Z_1^{t,u} \right] dt \\
&= -\frac{1}{\sqrt{2\pi}} \int_0^1 \frac{1}{1-t} \mathbb{E}\left[Z_1 e^{-\frac{Z_1^2}{2}} \mathbf{1}_{Z_1^{t,u} \geqslant 0}\, (\sqrt{t}\,u + \sqrt{1-t}\,Z_1) \right] dt,
\end{aligned}$$

since $\mathbb{E}[Z_2 \mathbf{1}_{Z_2^{t,u} \geqslant Z_1^{t,u}} \mid Z_1] = \frac{1}{\sqrt{2\pi}} \int_{Z_1}^{+\infty} z\, e^{-z^2/2} dz = \frac{1}{\sqrt{2\pi}} e^{-Z_1^2/2}$. Now,

$$
\begin{aligned}
&\int_0^1 \frac{1}{1-t} \mathbb{E}\left[Z_1 e^{-\frac{Z_1^2}{2}} \mathbf{1}_{Z_1^{t,u} \geqslant 0} \sqrt{t}\, u \right] dt \\
&= u \int_0^1 \frac{\sqrt{t}}{1-t} \mathbb{E}\left[Z_1 e^{-\frac{Z_1^2}{2}} \mathbf{1}_{Z_1 \geqslant -\sqrt{\frac{t}{1-t}} u} \right] dt \\
&= u \int_0^1 \frac{\sqrt{t}}{1-t} \int_{-\sqrt{\frac{t}{1-t}} u}^{+\infty} z e^{-z^2}\, dz\, dt \\
&= \frac{u}{2} \int_0^1 \frac{\sqrt{t}}{1-t} e^{-\frac{t}{1-t} u^2}\, dt \\
&= \frac{e^{u^2} u}{2} \int_0^{1/u^2} \frac{\sqrt{1-u^2\, t}}{t} e^{-\frac{1}{t}}\, dt.
\end{aligned}
$$

It is readily checked that the last integral is equivalent to $-2 \log u$, when $u \to 0^+$. On the other hand, by Fubini's theorem, we have

$$
\begin{aligned}
&\int_0^1 \frac{1}{\sqrt{1-t}} \mathbb{E}\left[Z_1^2 e^{-\frac{Z_1^2}{2}} (\mathbf{1}_{Z_1^{t,u} \geqslant 0} - \mathbf{1}_{Z_1^{t,0} \geqslant 0}) \right] dt \\
&= \int_0^1 \frac{1}{\sqrt{1-t}} \int_{-\sqrt{\frac{t}{1-t}} u}^{0} z^2 e^{-z^2} dz\, dt \\
&= \int_{-\infty}^0 z^2 e^{-z^2} \int_{\frac{z^2}{u^2+z^2}}^1 \frac{1}{\sqrt{1-t}} dt\, dz \\
&= \frac{u}{2} \int_{-\infty}^0 \frac{z^2}{\sqrt{u^2+z^2}} e^{-z^2}\, dz,
\end{aligned}
$$

which is a $\mathcal{O}(u)$ as $u \to 0^+$. This achieves the proof. □

Acknowledgements

We would like to thank Guillaume Carlier for suggesting the problem to us and subsequent useful discussions. We also thank Max Fathi for useful discussions. The three authors were supported by the Fonds de la Recherche Scientifique - FNRS under Grant MIS F.4539.16.

References

[1] A. D. Barbour. Stein's method for diffusion approximations. *Probability Theory and Related Fields*, 84(3):297–322, 1990.

[2] V. Bentkus. On the dependence of the Berry–Esseen bound on dimension. *Journal of Statistical Planning and Inference*, 113(2):385–402, 2003.

[3] A. Bikjalis. Estimates of the remainder term in the central limit theorem. *Litovsk. Mat. Sb*, 6:323–346, 1966.

[4] S. Bobkov. Entropic approach to E. Rio's central limit theorem for w2 transport distance. *Statistics & Probability Letters*, 83(7):1644–1648, 2013.

[5] S. Bobkov. Berry–Esseen bounds and Edgeworth expansions in the central limit theorem for transport distances. *Probability Theory and Related Fields*, 170(1-2):229–262, 2018.

[6] T. Bonis. Rates in the central limit theorem and diffusion approximation via Stein's method. *arxiv:1506.06966*, 2015.

[7] S. Chatterjee and E. Meckes. Multivariate normal approximation using exchangeable pairs. *Alea*, 4:257–283, 2008.

[8] L.HY Chen, L. Goldstein, and Q.-M. Shao. *Normal approximation by Stein's method*. Springer Science & Business Media, 2010.

[9] V. Chernozhukov, D. Chetverikov, and K. Kato. Central limit theorems and bootstrap in high dimensions. *The Annals of Probability*, 45(4):2309–2352, 2017.

[10] T. A. Courtade, M. Fathi, and A. Pananjady. Existence of Stein kernels under a spectral gap, and discrepancy bound. *Annales de l'Institut Henri Poincaré*, to appear 2018.
[11] M. Fathi. Stein kernels and moment maps. *arXiv:1804.04699*, 2018.
[12] R. E. Gaunt. Rates of convergence in normal approximation under moment conditions via new bounds on solutions of the Stein equation. *Journal of Theoretical Probability*, 29(1):231–247, 2016.
[13] F Götze. On the rate of convergence in the multivariate CLT. *The Annals of Probability*, 19(2):724–739, 1991.
[14] M. Ledoux, I. Nourdin, and G. Peccati. Stein's method, logarithmic Sobolev and transport inequalities. *Geometric and Functional Analysis*, 25(1):256–306, 2015.
[15] A. Müller. Integral probability metrics and their generating classes of functions. *Advances in Applied Probability*, 29(2):429–443, 1997.
[16] Y. S. Nefedova and I. G. Shevtsova. On nonuniform convergence rate estimates in the central limit theorem. *Theory of Probability & Its Applications*, 57(1):28–59, 2013.
[17] I. Nourdin and G. Peccati. *Normal approximations with Malliavin calculus: from Stein's method to universality*, volume 192. Cambridge University Press, 2012.
[18] M. Raič. A multivariate CLT for decomposable random vectors with finite second moments. *Journal of Theoretical Probability*, 17(3):573–603, 2004.
[19] G. Reinert and A. Röllin. Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. *The Annals of Probability*, 37(6):2150–2173, 2009.
[20] E. Rio. Upper bounds for minimal distances in the central limit theorem. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 45, pages 802–817. Institut Henri Poincaré, 2009.
[21] E. Rio. Asymptotic constants for minimal distance in the central limit theorem. *Electronic Communications in Probability*, 16:96–103, 2011.
[22] C. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory*. The Regents of the University of California, 1972.
[23] G. Valiant and P. Valiant. A CLT and tight lower bounds for estimating entropy. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 17, 2010.
[24] C. Villani. *Topics in Optimal Transportation*. Number 58. American Mathematical Soc., 2003.
[25] A. Zhai. A multivariate CLT in Wasserstein distance with near optimal convergence rate. *Probability Theory and Related Fields*, to appear.

Thomas O. Gallouët: Inria, Project team MOKAPLAN and Mathematics department, Université de Liège, Belgium, (thomas.gallouet@inria.fr)

Guillaume Mijoule: Inria, Project team MOKAPLAN and Mathematics department, Université de Liège, Belgium, (guillaume.mijoule@inria.fr)

Yvik Swan: Mathematics department, Université de Liège, Belgium, (yswan@ulg.ac.be)